AD 682953

הטכניון — מכון טכנולוגי לישראל

הפקולטה להנדסת תעשיה וניהול

TECHNION — ISRAEL INSTITUTE OF TECHNOLOGY

FACULTY OF INDUSTRIAL AND MANAGEMENT ENGINEERING

HAIFA, ISRAEL

19

A TIME-SHARING MODEL WITH MANY QUEUES

by

I. Adiri and B. Avi-Itzhak

Operations Research, Statistics and Economics

Mimeograph Series No. 20

This research has been sponsored in part by the Logistics and Mathematical Statistics Branch, Office of Naval Research, Washington D.C. under Contract F61052-68-C-0014.

July 1968

A B S T R A C T

The paper presents a mathematical study of a time-sharing system with a single server and many queues. In each queue a customer receives one quantum of service and is then sent to the end of the next queue, provided his service demand has not been completely satisfied. When a quantum of service is completed the server attends to the first customer in the lowest index non-empty queue. The discipline of the highest indexed queue is 'round-robin', while all other queues obey the FIFO rule. Newly arrived customers join the end of the first queue.

Mathematical expressions for the main performance measures are derived and illustrated by graphical means.

A TIME-SHARING MODEL WITH MANY QUEUES

I. Adiri and B. Avi-Itzhak

INTRODUCTION

One of the main advantages of time-sharing is the higher priority given to shorter service demands in cases where their exact length is not known in advance. In the conventional single queue time-sharing (round-robin) model (referred to as R.R.1 in Adiri and Avi-Itzhak [1]), service is given in small segments. Upon completion of a service segment, the customer is dismissed (i.e., either leaves the system or joins the end of the queue, depending upon whether his service has been completed or not) and the next in the queue is admitted. A newly arrived customer joins the end of the queue. In this system the server is ignorant of the amount of service (number of service segments - quanta) already delivered to the customers present in the system. Thus, it would be expected that the i-th response time of a customer (time between (i-1)-th and i-th admissions to service) is not strongly dependent upon i. It follows that the expected total response time (time in the system) of a customer is approximately an increasing linear function of his service-demand length. This is demonstrated in Figure 1, constructed on the basis of exact closed form results obtained in reference [1]. It relates to the steady state situation of an R.R.1 model where arrivals are assumed to behave as a homogeneous Poisson process with intensity λ, service demand lengths have a negative exponential distribution with mean $1/\mu$, and quantum length is $\tau + \Theta$, where τ is the set-up time and Θ the service segment (quantum processing time).

One weakness of time-sharing systems, inherent in the R.R.1 model, is sensitivity to overloading by customers with long service demands (long processing times). The presence of customers with long service demands increases the waiting times of all customers in the system, and in cases of overloading (saturation) a demand cannot be satisfied in a finite time, however short it may be. On the other hand, there is no sufficient discouragement of customers with long service demands, since waiting time increases linearly with demand length. An actual time-sharing

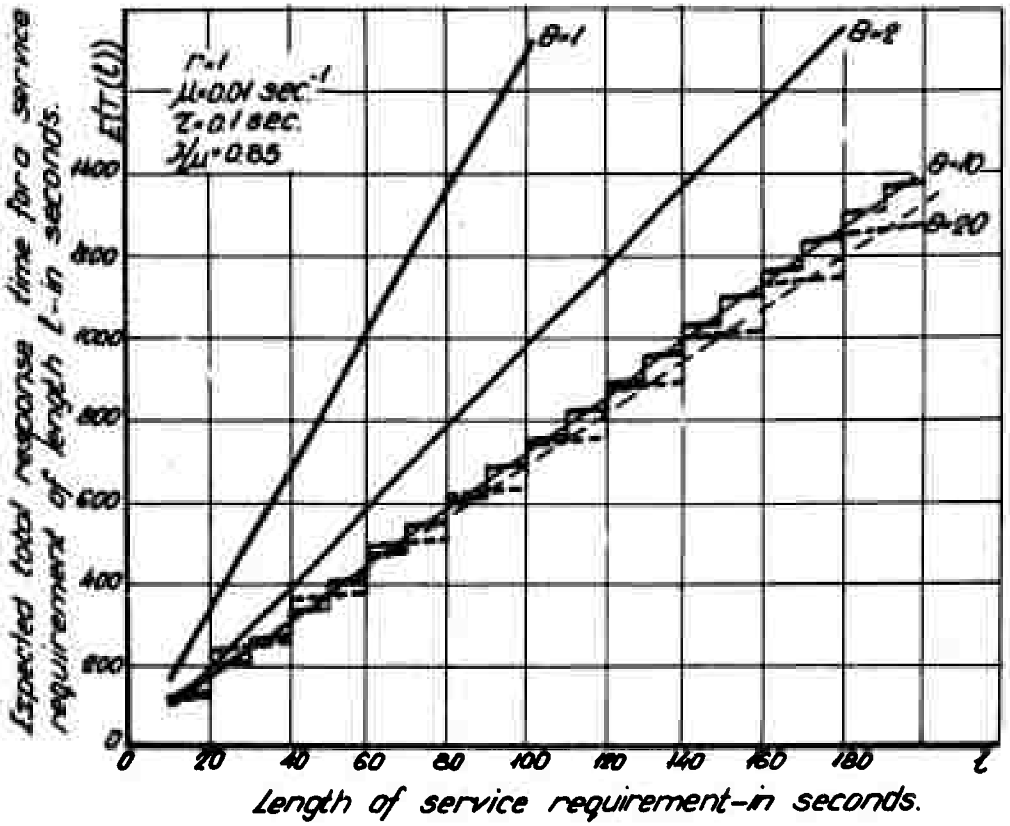

Figure 1: Expected Total Response Time for a Service Requirement of Length l as a Function of l (For different values of θ, The R.R.1 Model)

system is frequently overloaded during peak hours and underexploited during the rest of the day. It is desirable to automatically transfer long service demands, placed during peak hours, to less loaded hours of the day, or even discourage them completely and force them to use batch processing during night hours. What is needed is, thus, a system with the following features: (i) waiting times of short demands are less sensitive to the presence of long demands than in the simple round-robin discipline; (ii) short demands satisfied even under conditions of overloading (the higher the overloading, the shorter the demand must be in order to be satisfied); and (iii) effective discouragement of long demands during peak hours, by postponing their service to non-peak hours. Needless to say, the above must be realised without prior knowledge of demand lengths.

The R.R.r. model is proposed as a solution along the above lines. It comprises one server and r queues all of which, except the r-th, observe the FIFO rule, while the r-th queue is operated as an R.R.1. A newly arrived customer joins the end of the first queue where, in due course, he receives a certain amount of service (processing), following which he either departs or joins the end of the second queue for additional service. The server, after completing a service quantum, selects the first customer in the lowest index (highest priority) queue. The limit case where $r = \infty$, i.e. R.R.∞, has been first suggested and studied by Schrage [2].

The number of queues, r, and the quantum processing time for each queue, θ, may serve as control parameters of the system.

MATHEMATICAL MODEL - R.R.r.

Customers arrive at a single server facility according to a homogeneous Poisson process with intensity λ. Service times (as demanded by customers) are independent random variables identically distributed exponentially with mean $1/\mu$. The waiting line consists of r separate queues. A newly arrived customer joins the end of the first queue; upon admittance to service he is eligible for θ_1 units of service time. If his service demand is shorter than θ_1, he completes his service and leaves the system, otherwise he receives θ_1 units of service time and joins the end of the second queue. In general, at the i-th queue $(i=1,2,\ldots,r-1)$ the customer is eligible for θ_i units of service time during which he may complete his service and leave the system, or join the end of the (i+1)-th queue after receiving θ_i units of service time. A customer in the r-th queue who does not complete his

service within Θ_r service time units re-joins the end of the same queue. Having completed a service quantum at any one of the queues, the server admits to service the customer who is first in the lowest index non-empty queue. We assume that each quantum starts with a set-up time due to swapping, housekeeping etc. The i-th quantum thus comprises two elements: a set-up time of length τ_i and a processing time not exceeding Θ_i units.

Busy Period and Backlog

Let:

$$\alpha_i = e^{-\mu\Theta_i} , \tag{1}$$

$$\beta_i = e^{-\mu\sum_{j=1}^{i}\Theta_j} = \prod_{j=1}^{i} \alpha_j \tag{2}$$

$$q_i = \sum_{j=1}^{i}(\Theta_j + \tau_j) , \qquad i=1,2,\ldots, \tag{3}$$

where $(1-\alpha_i)$ is the probability of service being completed in the i-th queue, $i \leq r$ For $i > r$, Θ_i and τ_i are defined as equal to Θ_r and τ_r, respectively. Hence for $i > r$ we have $\alpha_i = \alpha_r$, $\beta_i = \beta_r$ and $q_i = q_r + (i-r)(\Theta_r+\tau_r)$.

The length of a quantum devoted to a customer in the i-th $(i=1,2,\ldots,r)$ queue behaves as a random variable which we shall denote by Q_i.

$$Q_i \overset{d}{=} \begin{cases} \Theta_i + \tau_i & \text{with probability } \alpha_i \\ D_i & \text{with probability } 1-\alpha_i \end{cases} \tag{4}$$

where D_i is a r.v. possessing the density:

$$f_{D_i}(x) = \frac{\mu e^{-\mu(x-\tau_i)}}{1 - \alpha_i} , \qquad \tau_i \leq x \leq \Theta_i + \tau_i, \quad i=1,2,\ldots,r. \tag{5}$$

For $i > r$, $Q_i \overset{\text{def}}{=} Q_r$ and $D_i \overset{\text{def}}{=} D_r$.

Within each queue, quantum lengths are mutually independent and also independent of the number of customers present in the various queues. This property is general for all time distributions, except for the r-th queue where it may not hold for non-exponential service times. The quanta in the different queues are mutually independent

Let the r.v. S denote the total time (including set-ups) devoted by the server to a customer.

$$S \overset{d}{=} \begin{cases} D_1 & \text{with probability } 1 - \alpha_1 \\ q_i + D_{i+1} & \text{with probability } \beta_i(1 - \alpha_{i+1}) \\ & i=1,2,\ldots, \end{cases} \tag{6}$$

The backlog and the busy period of the R.R.r. model behave as in a M/G/1 model with service times distributed as S. This equivalence holds also for R.R.1 systems with general time distributions, where S is defined as the total time devoted to a customer.

Total Response Time

The total response time of a customer is defined as the total time he spends in the system from arrival to departure.

Let X_n $(n=1,2,\ldots,)$ denote the number of customers in the R.R.r system at the moment of the n-th departure. Consider now a similar R.R.∞ model with the same system parameters except that for $i > r$, $\Theta_i = \Theta_r$ and $\tau_i = \tau_r$, and let X'_n denote the number of customers in the R.R.∞ system at the moment of the n-th departure. Clearly, X_n and X'_n are equal in distribution. This is a result of the premise that service times are exponentially distributed, and is explained as follows: Up to and including the (r-1)-th queue, both systems are identical; the input processes into their respective r-th queues are also identical, and the departure process from the r-th queue of the R.R.r model is statistically identical to that from queues r, r+1,..., of the R.R.∞ model. According to Little's Theorem [3], the steady state expectation of total response time (denoted by E(T)), is identical in both systems.

Using Schrage's results [2] for the R.R.∞ model, we obtain:

$$E(T) = \int_0^\infty E(T|L=\ell)\, \mu e^{-\mu\ell} d\ell =$$

$$= \frac{1}{\mu} + \sum_{n=1}^{\infty} \beta_{n-1}(1 - \alpha_n)\left(\frac{\lambda E(U_n^2) + \lambda \sum_{i=n+1}^{\infty} \beta_{i-1} E(Q_i^2)}{2(1 - \lambda E(U_{n-1}))(1 - \lambda E(U_n))} +\right.$$

$$\left. + \frac{q_{n-1}}{1 - \lambda E(U_{n-1})} - \sum_{i=1}^{n-1} \Theta_i + \tau_n\right) \quad . \tag{7}$$

In this expression $E(T|L=\ell)$ is the total response time expectation of a customer who demands ℓ service time units in the R.R.∞ model, $\alpha_0=\beta_0=1$, $q_0=\Theta_0=\tau_0=0$, and

$$U_n \overset{d}{=} \begin{cases} q_{i-1} + D_i & \text{with probability } \beta_{i-1}(1-\alpha_i), \quad i=1,2,\dots,n \\ q_n & \text{with probability } \beta_n \quad , \end{cases} \qquad n=1,2,\dots, \tag{8}$$

$$E(U_n) = q_n\beta_n + \sum_{i=1}^{n} \beta_{i-1}(1-\alpha_i)(q_{i-1} + E(D_i))$$

$$E(U_n^2) = q_n^2\beta_n + \sum_{i=1}^{n} \beta_{i-1}(1-\alpha_i)(q_{i-1}^2 + 2q_{i-1}\, E(D_i) + E(D_i^2)) \qquad n=1,2,\dots, \tag{9}$$

Conditional Total Response Time - R.R.r

The total steady state response time of a customer with service requirement ℓ is denoted by $T(\ell)$. This quantity will be referred to as the conditional total response time.

The behavior of the first $(r-1)$ queues is the same in both the R.R.r and the R.R.∞ systems. Thus for $\ell \leq \sum_{i=1}^{r-1} \Theta_i$ we may use Schrage's results [2].

$$E(T(\ell)) = \frac{\lambda E(U_n^2) + \lambda_i \sum_{i=n+1}^{\infty} \beta_{i-1} E(Q_i^2)}{2(1 - \lambda E(U_{n-1}))(1 - \lambda E(U_n))} + \frac{q_{n-1}}{1 - \lambda E(U_{n-1})} + \ell - \sum_{i=1}^{n-1} \Theta_i + \tau_n \,, \tag{10}$$

where n is an integer such that

$$\sum_{i=1}^{n-1} \Theta_i < \ell \leq \sum_{i=1}^{n} \Theta_i \quad . \tag{11}$$

Customers with service time $\ell > \sum_{i=1}^{r-1} \Theta_i$ reach the r-th queue, where they circulate until departure. We define the i-th $(i=2,3,\dots,)$ response time of a customer (denoted by T_i) as the time elapsing between his (i-1)-th and (i-th) admissions to service, assuming that $\ell > \sum_{j=1}^{i-1} \Theta_j$. Similarly, T_1 is the time elapsing between his arrival and first admission. Let (K_i+1) denote the number of customers present at the time of the i-th admission of a customer to service. For $i=r,r+1,\dots,$ (K_i+1) is their number in the r-th queue at the moment of an i-th admission. The i-th response time expectation, $E(T_i)$, $(i=r+1,r+2,\dots,)$ is obtained as:

$$E(T_i) = \frac{\theta_r + \tau_r + E(K_{i-1})E(Q_r)}{1 - \lambda E(U_{r-1})} , \qquad i=r+1,r+2,\ldots, \tag{12}$$

where $r \geq 2$.

The argument supporting Eq. (12) is that T_i may be viewed as a busy period starting with initial expected backlog of magnitude $(\theta_r + \tau_r + E(K_{i-1})E(Q_r))$ in a M/G/1 system where the arrival rate is λ and service times are distributed as U_{r-1}.

For $\sum_{i=1}^{n-1} \theta_i < \ell \leq \sum_{i=1}^{n} \theta_i$ we have:

$$E(T(\ell)) = \sum_{i=1}^{n} E(T_i) + \ell - \sum_{i=1}^{n-1} \theta_i + \tau_n \quad . \tag{13}$$

Substitution of Eq. (12) yields for $n \geq r$:

$$E(T(\ell)) = \sum_{i=1}^{r} E(T_i) + \frac{(n-r)(\theta_r + \tau_r) + E(Q_r)\sum_{i=r}^{n-1} E(K_i)}{1 - \lambda E(U_{r-1})} + \\ + \ell - \sum_{i=1}^{n-1} \theta_i + \tau_r \quad . \tag{14}$$

In this expression both sums are unknown. We sould like to express them as functions of $E(K_r)$. The following recursive relation is readily shown to hold for $i > r$,

$$E(K_i) = E(T_i)\lambda\beta_{r-1} + E(K_{i-1})\alpha_r \quad . \tag{15}$$

From Eqs. (12) and (15) we obtain:

$$\sum_{i=r}^{n-1} E(K_i) = (E(K_r) - \frac{a}{1-b}) \frac{1-b^{n-r}}{1-b} + \frac{(n-r)a}{1-b} \quad , \tag{16}$$

where:

$$a = \frac{\lambda\beta_{r-1}(\theta_r + \tau_r)}{1 - \lambda E(U_{r-1})} \tag{17}$$

and:

$$b = \alpha_r + \frac{\lambda\beta_{r-1}E(Q_r)}{1 - \lambda E(U_{r-1})} \quad . \tag{18}$$

Let T_c be the time elapsing between the arrival of a customer whose service time exceeds $\sum_{i=1}^{r-1} \Theta_i$ and next admission of a customer in the r-th queue; and let Z denote the number of customers in the r-th queue when a new customer joins it. The time elapsing between admission of a customer from the r-th queue until the server is again ready to admit a customer from it behaves as a busy period of a M/G/1 model, where service times are distributed as U_{r-1} and the starting backlog is distributed as Q_r. Bearing in mind the definitions of T_c and Z, we thus have:

$$\sum_{i=1}^{r} E(T_i) = E(T_c) + \frac{E(Z)E(Q_r)}{1 - \lambda E(U_{r-1})} . \tag{19}$$

The expectation of T_c is obtained by using Schrage's [2] approach.

$$E(T_c) = \frac{\lambda E(U_{r-1}^2) + \lambda E(Q_r^2)\beta_{r-1}/(1 - \alpha_r)}{2(1 - \lambda E(U_{r-1}))^2} + \frac{q_{r-1}}{1 - \lambda E(U_{r-1})} . \tag{20}$$

Since arrivals and departures take place one at a time, the number left in the r-th queue by a departing customer must be distributed as Z. (This is a steady state argument). Let χ be the number present in the r-th queue at the moment when a customer of this queue is admitted to service, then

$$E(Z) = E(\chi) - 1 , \qquad \text{for } r \geq 2 . \tag{21}$$

On the other hand, we have:

$$E(\chi) = (1 - \alpha_r) \sum_{i=r}^{\infty} E(K_i)\alpha_r^{i-r} + 1 \tag{22}$$

Substituting of Eqs. (12), (15) and (21) in (22), summation and substitution in (21) yield:

$$E(Z) = \frac{\alpha_r a}{1 - \alpha_r b} + \frac{1 - \alpha_r}{1 - \alpha_r b} E(K_r) . \tag{23}$$

Substituting Eqs. (20) and (23) in (19), we obtain:

$$\sum_{i=1}^{r} E(T_i) = \frac{\lambda E(U_{r-1}^2) + \lambda E(Q_r^2)\beta_{r-1}/(1-\alpha_r)}{2(1-\lambda E(U_{r-1}))^2} + \frac{q_{r-1}}{1-\lambda E(U_{r-1})} +$$

$$+ \left(\frac{\alpha_r a}{1-\alpha_r b} + \frac{1-\alpha_r}{1-\alpha_r b} E(K_r)\right) \frac{E(Q_r)}{1-\lambda E(U_{r-1})} \quad . \tag{24}$$

Substituting Eqs. (16) and (24) in (14), we have $E(T(\ell))$ expressed as a function of $E(K_r)$ and of the various parameters of the process.

$$\begin{aligned} E(T(\ell)) = {} & \frac{\lambda E(U_{r-1}^2) + \lambda E(Q_r^2)\beta_{r-1}/(1-\alpha_r)}{2(1-\lambda E(U_{r-1}))^2} + \frac{q_{r-1}}{1-\lambda E(U_{r-1})} + \\ & + \frac{b-\alpha_r}{\lambda\beta_{r-1}} \Big\{ (\alpha_r a + (1-\alpha_r)E(K_r))/(1-\alpha_r b) + \\ & + (E(K_r) - a/(1-b))(1-b^{n-r})/(1-b) + (n-r)a/(1-b) \Big\} \\ & + (n-r)a/(\lambda\beta_{r-1}) + \ell - \sum_{i=1}^{n-1} \theta_i + \tau_r \quad . \end{aligned} \tag{25}$$

$E(K_r)$ is obtainable from:

$$E(T) = \int_0^\infty E(T(\ell))\, \mu e^{-\mu\ell} d\ell \quad , \tag{26}$$

where $E(T)$ being given in Eq. (7), $E(T(\ell))$ for $n < r$ by Eq. (10) and for $n \geq r$ by Eq. (25). The result obtained for $E(K_r)$ is fairly complicated in form:

$$\begin{aligned} E(K_r) = {} & \left\{ \frac{b-\alpha_r}{\lambda\beta_{r-1}} \left(\frac{1}{1-\alpha_r b} + \frac{1}{1-b}\left(\frac{1}{1-\alpha_r} - \frac{1}{1-\alpha_r b}\right)\right)\right\}^{-1} \\ & \left\{ \sum_{n=r}^{\infty} \alpha^{n-r} \left(\frac{\lambda E(U_n^2) + \lambda E(Q_r^2)\beta_{r-1}\alpha_r^{n+1-r}/(1-\alpha_r)}{2(1-\lambda E(U_{n-1}))(1-\lambda E(U_n))} + \frac{q_{n-1}}{1-\lambda E(U_{n-1})}\right) - \right. \\ & - \frac{\alpha_r}{(1-\alpha_r)^2} \left(\frac{a}{\lambda\beta_{r-1}} \left(1 + \frac{b-\alpha_r}{1-b}\right)\right) - \frac{1}{1-\alpha_r} \left[\frac{\lambda E(U_{r-1}^2) + \lambda E(Q_r^2)\beta_{r-1}/(1-\alpha_r)}{2(1-\lambda E(U_{r-1}))^2} + \right. \\ & \left.\left. + \frac{q_{r-1}}{1-\lambda E(U_{r-1})} + \frac{b-\alpha_r}{\lambda\beta_{r-1}} \left(\frac{\alpha_r a}{1-\alpha_r b} - \frac{a}{(1-b)^2} \left(1 - \frac{(1-\alpha_r)}{(1-\alpha_r b)}\right)\right)\right]\right\} . \end{aligned} \tag{27}$$

DISCUSSION

The complexity of the mathematical formulae rules out direct analysis and discussion of the results, and therefore the authors have decided to illustrate the main features of the model by means of graphs.

Figures 2 and 3 show, for different values of r, the functional relation between the total response time of a customer and the length of his service time. As r increases waiting of short demands decreases while waiting of long demands increases. In these examples $\Theta_i = \Theta$, $i=1,2,\ldots,r$, hence the unconditional expected total response time is independent of r. Obviously, any decrease in waiting time of short demands is obtained at the expense of the long demands. The terms "short demand" and "long demand" are used here qualitatively. In a real situation it suffices to define an overall objective function and optimize with respect to the control parameters r and Θ. It should be noted that the numerical results for $r = 1$ were obtained by using the mathematical results of reference [1], since the R.R.r model treated in this paper does not include the case $r = 1$.

Another way of increasing priority to short customers is by increasing the magnitude of Θ, (assuming $\Theta_i = \Theta$, $i=1,2,\ldots,r$). This is illustrated in Figures 4 and 5 where r is kept constant for different values of Θ. Note that when Θ exceeds a certain magnitude priority to short services is reduced and the queue discipline tends to FIFO. In Figure 5 the load on the station is relatively low ($\lambda/\mu = 0.5$). Consequently, the effect of increasing Θ is not as noticeable as in Figure 4 ($\lambda/\mu = 0.85$). In both figures the time losses, represented by set-ups, decrease when Θ is increased, since the value of τ (assuming $\tau_i = \tau$, $i=1,2,\ldots,r$) is kept constant.

In the preceding examples identical values of Θ_i were used in all queues, i.e., $\Theta_i = \Theta$, $i=1,2,\ldots,r$. The priority given to short demands may be increased at the expense of the long demands, by using a decreasing sequence (Θ_i, $i=1,2,\ldots,r$) of quantum processing times. Here an R.R.r system with a fixed quantum processing time (equaling Θ in all queues) is concerned with equivalent system in which Θ_i is an arbitrary sequence. It is assumed that set-up times are the same constant for

all queues in both systems, i.e., $\tau_i = \tau$, $i=1,2,\ldots,r$, and furthermore, that all other parameters, except for the quantum processing times, are identical in both systems. For the comparison to be meaningful, the expected time losses (represented by set-ups) per service must be equal in both systems. This implies that the expected number of quanta per service is the same in both. Expressing this equivalence mathematically, we have:

$$\sum_{i=1}^{\infty} i(e^{-\mu\theta})^{i-1}(1 - e^{-\mu\theta}) = \sum_{i=1}^{r-1} i\beta_{i-1}(1 - \alpha_i) + \beta_{r-1}(1 - \alpha_r) \sum_{i=r}^{\infty} i\alpha_r^{i-r} \quad . \tag{28}$$

Summation yields:

$$\frac{1}{1 - e^{-\mu\theta}} = \sum_{i=1}^{r} i\beta_{i-1}(1 - \alpha_i) + r\beta_{r-1} + \frac{\beta_{r-1}\alpha_r}{1 - \alpha_r} \quad . \tag{29}$$

For any given sequence $(\theta_i,\ i=1,2,\ldots,r)$ the value of θ can be uniquely determined using Eq. (29).

Figure 6 illustrates an example where the sequence $(\theta_i,\ i=1,2,\ldots,r)$ is of the form $\theta_{i+1} = 0.95\theta_i$, $i=1,2,\ldots,r-1$ and $\theta_1 = 10$, $r = 50$. The equivalent value of θ for the system where $\theta_i = \theta$, $i=1,2,\ldots,r$ is $\theta = 2.81$.

As mentioned in the introductionm during peak-hours (oversaturation) the R.R.r system still satisfies short demands, while service to long demands is automatically postponed to non-peak hours. This is shown graphically in Figure 7. Table 1 supplements Figure 7 by giving the longest demand that can be still served in a finite time for each traffic intensity.

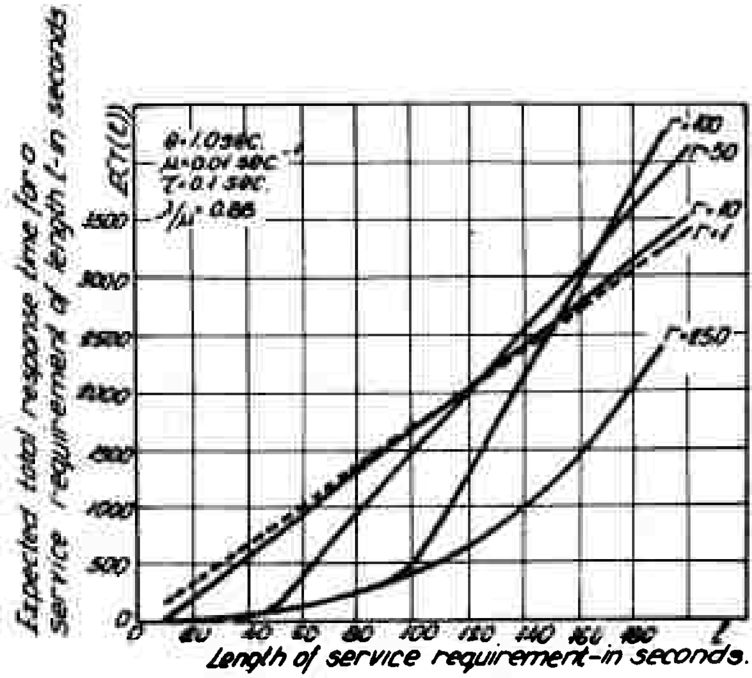

Figure 2: Expected Total Response Time for a Service Requirement of Length l as a Function of l. (For different values of r.)

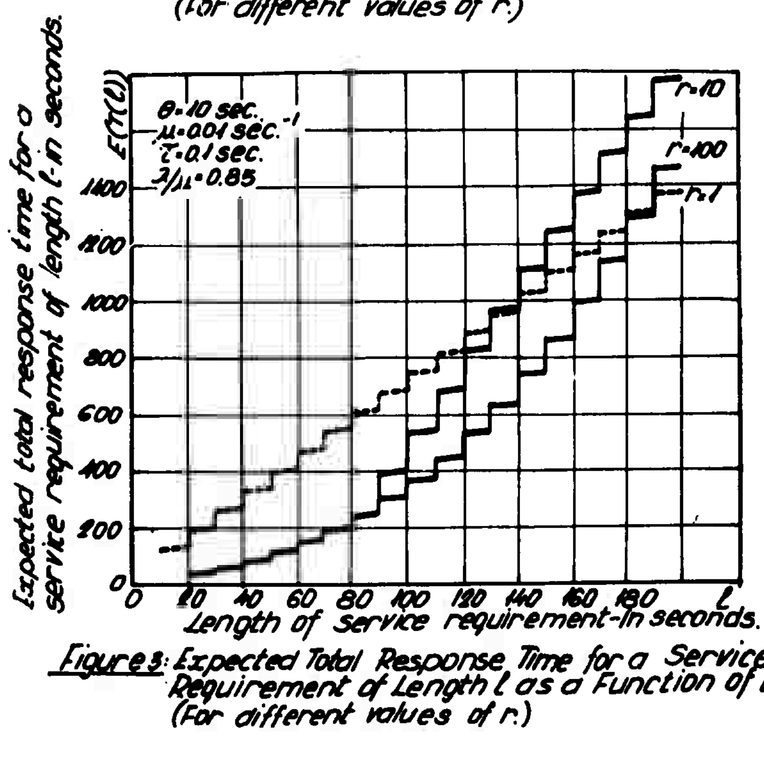

Figure 3: Expected Total Response Time for a Service Requirement of Length l as a Function of l. (For different values of r.)

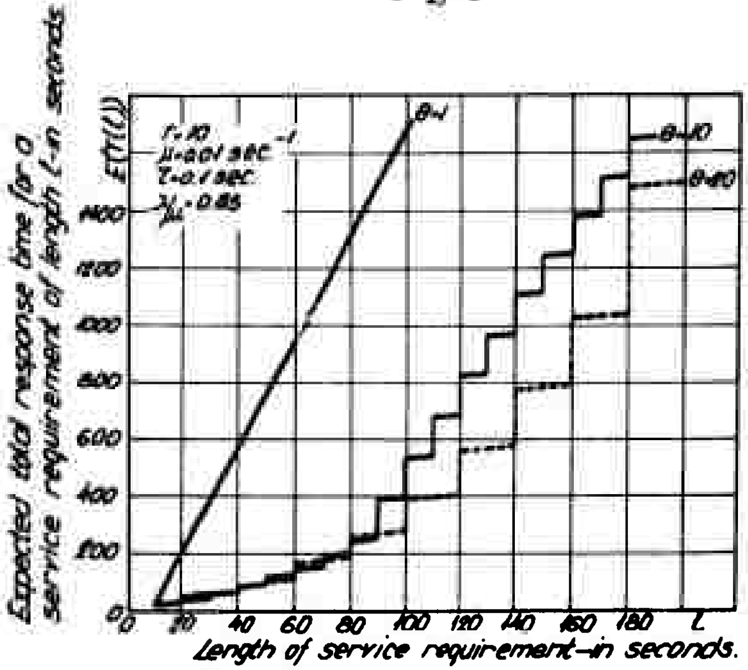

Figure 4: Expected Total Response Time for a Service Requirement of Length l as a Function of l. (For different values of θ)

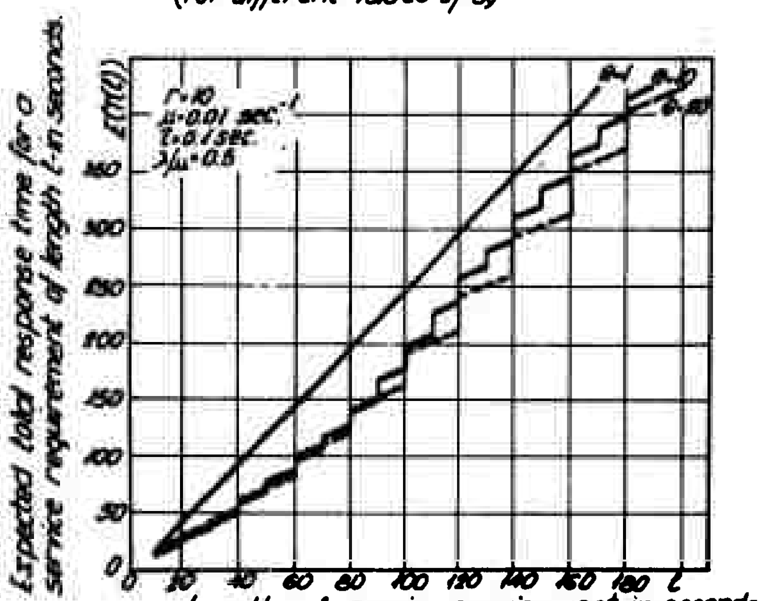

Figure 5: Expected Total Response Time for a Service Requirement of Length l as a Function of l. (For different values of θ.)

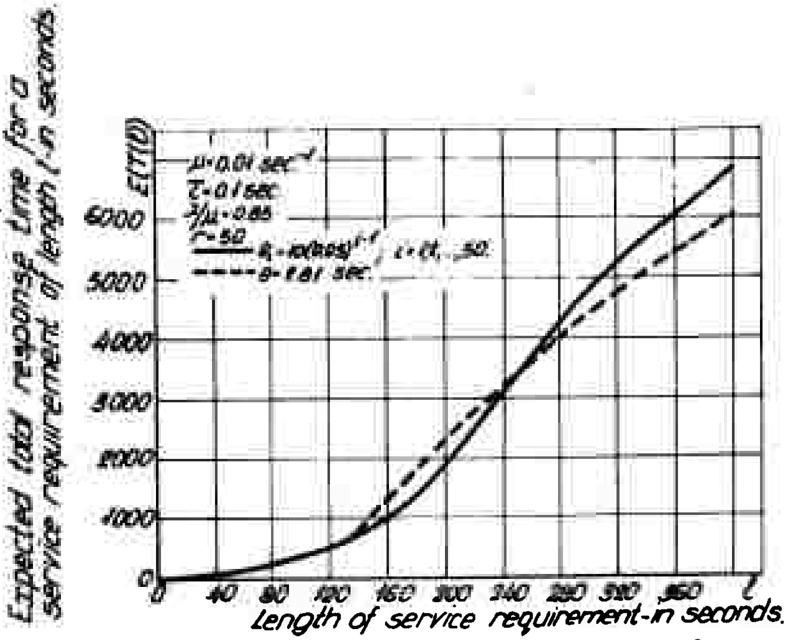

Figure 6: Expected Total Response Time for a Service Requirement of Length l as a Function of l.

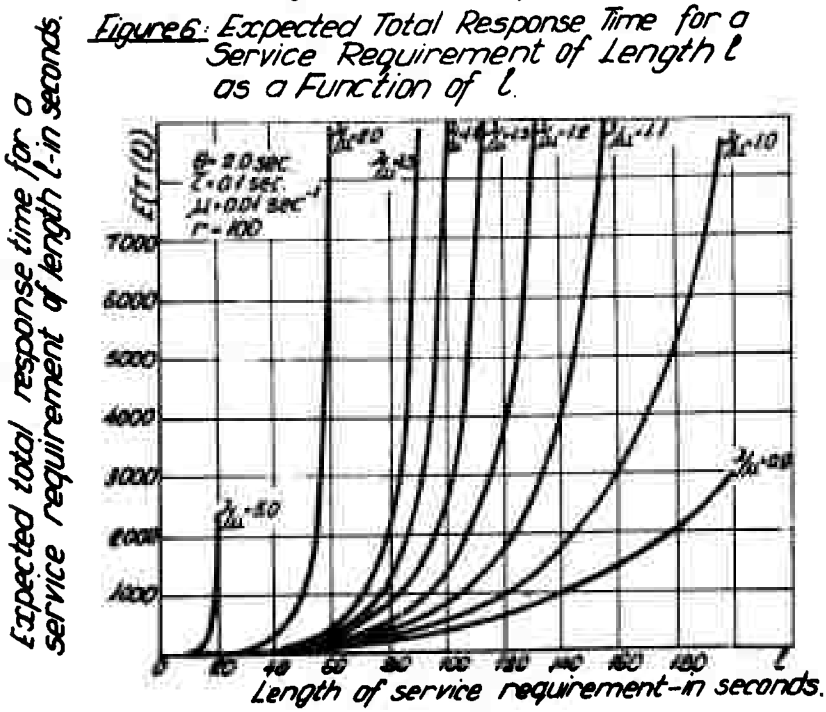

Figure 7: Expected Total Response Time for a Service Requirement of Length l as a Function of l
(For different values of trafic intensity)

Table 1: Largest Satisfied Demand as a Function of Traffic Intensity.
(r = 100, μ = 0.01 sec.$^{-1}$, τ_i = 0.1 sec, θ_i = 2 sec, $i=1,2,\ldots,r$)

λ/μ	$\rho = \lambda E(S)$*	number of unsaturated queues	largest satisfied demand in seconds
0.9	0.95	100	all demands
1.0	1.05	99	198
1.1	1.16	99	198
1.2	1.26	78	156
1.3	1.37	65	130
1.4	1.47	56	112
1.5	1.58	50	100
2.0	2.10	32	64
5.0	5.25	10	20

*Note that the steady state condition is $\rho < 1$.

REFERENCES

1. Adiri, I. and B. Avi-Itzhak, "A time-sharing queue", Opes. Re.. Stat. and Econ., Mimeograph Series No. 5, Fac. of Ind. and Mgt. Eng., Technion. Submitted for publication.

2. Schrage, L.E., "The Queue M/G/1 with Feedback to Lower Priority Queues", Management Science, Vol. 13, No. 7, March 1967, pp. 466-474.

3. Little, J.D.C., "A Proof for the Queueing Formula $L = \lambda W$", Operations Research. Vol. 9, No. 1, 1961, pp. 70-76.

O.R. Mimeograph Series Publications

1.	Mitrani, I. & Avi-Itzhak, B.	"A Many Server System with Service Interruptions".
2.	Haitovsky, Y.	"Missing Data in Regression Analysis".
3.	Haitovsky, Y.	"Efficient least squares regression of differently grouped observations when the cross classifications are unknown".
4.	Powell, B.A. & Avi-Itzhak, B.	"Queueing Systems with Enforced Idle Time".
5.	Adiri, I. & Avi-Itzhak, B.	"A Time-Sharing Queue".
6.	Levin, O.	"Optimal Control of a Storage Reservoir During a Flood Season".
7.	Adiri, I.	"A Time-Sharing Queue with Preemptive Resume Priority Discipline".
8.	Haitovsky, Y.	"The Theory of Regression Estimation from Grouped Observations".
9.	Pollatschek, M.A. & Avi-Itzhak, B.	"Algorithms for Stochastic Games with Geometrical Interpretation".
10.	Adiri, I. & Avi-Itzhak, B.	"A Time-Sharing Queue with a Finite Number of Customers".
11.	Haitovsky, Y.	"A Note on the Maximization of $\bar{R}^2$".
12.	Naor, P.	"On the Regulation of Queue Size by Levying Tolls".
13.	Adiri, I.	"Computer Time-Sharing Queue with Priorities".
14.	Avi-Itzhak, B. & Mandelbaum, M.	"Mathematical Models for Outgoing Traffic Flow in an Airport Termina
15.	Avi-Itzhak, B. & Mandelbaum, M.	"Introduction to Queueing with Splitting and Matching".
16.	Pollatschek, M.A. & Levin, O.	"On the Optimization Model of Groundwater Utilization".
17.	Kondor, Y.	"Linear Programming of Income Tax Rates".
18.	Hammer, P.L.	"Increasing the Capacity of a Network".
19.	Hammer, P.L.	"Time Minimizing Transportation Problems".
20.	Adiri, I. & Avi-Itzhak, B.	"A Time-Sharing Model with Many Queues".